4K

УДК 004.032.26

А.В. Кондратюк, К.В. Чумичкин
Одесский национальный политехнический университет, г. Одесса, Украина
alkond@ukr.net, neuroschool@narod.ru

Показатель чувствительности нейронной сети к входным данным

В статье исследуется проблема потери информативности входных данных на этапе синтеза искусственной нейронной сети при решении задач прогнозирования временных рядов. Предложен подход к решению данной проблемы, основанный на повышении чувствительности нейронной сети к входным данным путем анализа их статистических характеристик. Предложено решение актуальной задачи разработки показателя, определяющего степень чувствительности нейронной сети к значениям входных данных. Благодаря разработке такого показателя стало возможным определить виды входных данных, наиболее эффективно обрабатываемые нейронной сетью. Эффективность предложенного показателя обоснована и подтверждена теоретическими выкладками и проиллюстрирована наглядно на рисунках. Обозначены перспективы дальнейшего развития данного направления исследований.

Введение

В настоящее время всё более актуальными становятся задачи оптимизации, поиска, реализации распределенных и параллельных систем, а также задачи прогнозирования. Многие из них при реализации математическими методами весьма трудоемки, а порой и вовсе не разрешимы и, следовательно, требуют к себе «особых» подходов, обладающих большей универсальностью и требующих менее детальной формализации. К числу таких подходов относятся нейросетевые методы обработки информации, которые нашли широкое применение в различных предметных областях. Такие преимущества нейросетевых методов, как слабая чувствительность к шуму и пропускам в данных, возможность параллельной обработки данных высокой размерности, использование «обучения» вместо трудоемкой алгоритмизации задач, и сделали их наиболее популярными, особенно при решении нетривиальных задач интеллектуального анализа данных.

Данная работа рассматривает входные данные, представленные в виде временного ряда (ВР), т.е. совокупности значений изменения некоторой величины с течением времени, и многослойные искусственные нейронные сети (ИНС) типа «многослойный персептрон», обучаемые с учителем по классическому алгоритму обучения (алгоритм обратного распространения ошибки) [1], [2]. Многослойный персептрон является универсальным средством аппроксимации, как показано в [3], а также обладает устойчивым поведением.

В процессе синтеза ИНС существует три основных этапа:

1 – формирование обучающей выборки (ОВ);

2 – структурный синтез ИНС;

3 – параметрический синтез (обучение) ИНС.

Одним из наиболее важных этапов в решении задачи нейросетевого прогнозирования является формирование ОВ. От ее состава, полноты и качества в значительной мере зависят время обучения ИНС и достоверность получаемых

моделей. Поэтому, как правило, на данном этапе синтеза ИНС имеет место целый ряд различных предварительных преобразований, направленных на предобработку «сырых» исходных данных, а именно: восстановление пропущенных значений, вычитание шума, устранение незначащих факторов, сглаживание, фильтрация данных, расчет различных показателей, сортировки, группировки и ряд других операций [4].

Однако при обучении ИНС может иметь место потеря точности восстановления функциональных зависимостей, которая вызвана неравномерной крутизной активационной функции нейрона (АФН) в диапазоне допустимых входных значений [5].

Кроме того, для большинства ИНС характерно наличие диапазона допустимых значений входных данных, в пределах которого они различимы [6]. АФН входного слоя ИНС определяет допустимые границы значений входных данных.

Как правило, в допустимом диапазоне присутствуют как слабо заполненные участки, так и участки скученности значений входной величины. Слабо заполненные участки приводят к тому, что в процессе обучения ИНС плохо «запоминает» эти значения. А участки скученности, где на относительно небольших отрезках располагается значительное количество значений входной величины, оказываются слабо различимыми для ИНС.

При этом возникает ситуация, когда ИНС воспринимает группу очень близко расположенных значений входной величины как одно значение, что соответственно приводит к понижению информационной энтропии входов ИНС. Это, в свою очередь, отрицательно сказывается на информативности входных данных, приводя к частичной их потере [5].

В то же время путем подстройки параметров данной АФН к особенностям распределения значений исследуемого ВР можно добиться учета его статистических характеристик и не допустить потерь в точности вычислений.

Таким образом, возникает проблема учета статистических характеристик входных данных. Решением этой проблемы и целью данной работы является разработка показателя, определяющего степень чувствительности ИНС к значениям входных данных, т.е. уровень их различимости с точки зрения ИНС.

Поскольку статистическая чувствительность ИНС к данным определяется типом АФН и её параметрами, то задачей настоящей работы является сокращение затрат времени на синтез ИНС путем разработки показателя чувствительности ИНС к статистическим характеристикам входных данных, что позволит заблаговременно, т.е. еще до этапа параметрического синтеза ИНС, оптимальным образом настроить параметры АФН и получить более быструю и адаптированную к специфике входных данных информационную систему интеллектуального анализа данных.

Показатель чувствительности нейронной сети

Как было отмечено выше, чувствительность ИНС к статистическим характеристикам входных данных определяет АФН.

Рассмотрим идеализированную ситуацию, когда входные данные нормированы и распределены по нормальному закону распределения $F_1(x)$ и в качестве АФН входного слоя ИНС используется нормированная сигмоидальная АФН $F_2(x)$ [7].

Под нормированием входных данных в данном случае понимается отображение значений входных данных в некоторый допустимый диапазон [a, b], что на практике осуществляется с помощью предварительных преобразований на

этапе формирования ОВ (рис. 1). Пусть входные данные имеют нормальный закон распределения (1). Известно, что некоторая случайная величина X распределена по нормальному закону (закону Гаусса) с параметрами m, σ, если

$$f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \; x \in R \tag{1}$$

В случае, если m = 0, σ = 1, распределение называется нормированным:

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{2}$$

Под нормированной понимается такая АФН, у которой параметры подобраны таким образом, что область значений функции соответствует допустимому диапазону входных данных [a, b] (рис. 1).

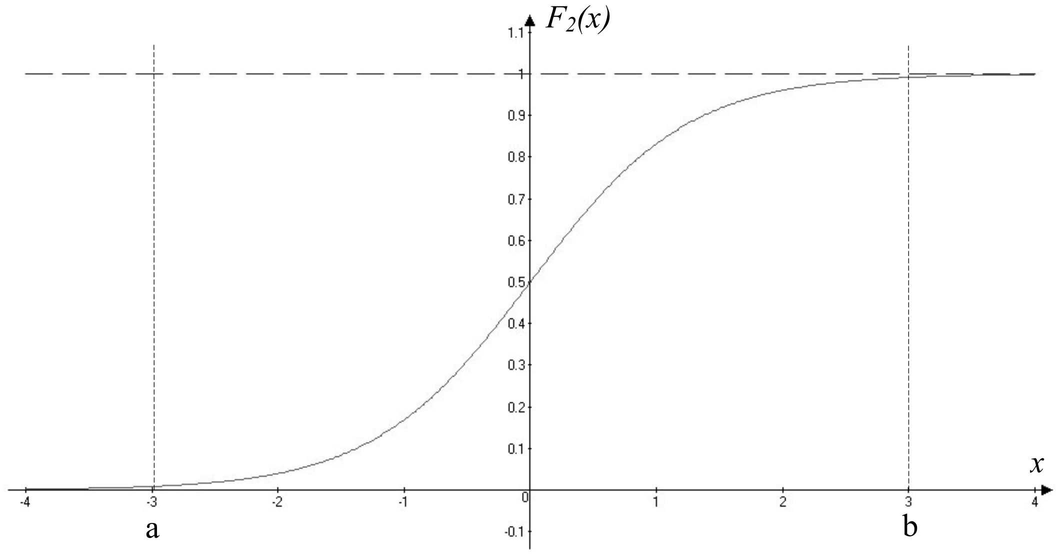


Рисунок 1 – График нормированной сигмоидальной АФН

Экспериментально подобрав значения параметра α сигмоидальной функции

$$F(x) = \frac{1}{1+e^{-ax}}, \tag{3}$$

где α – коэффициент, определяющий кривизну функции, получим нормированную сигмоидальную АФН:

$$F_2(x) = \frac{1}{1+e^{-1,6x}}. \tag{4}$$

Первая производная функции (4) будет иметь вид:

$$f_2'(x) = \frac{1,6\, e^{-1,6x}}{(1+e^{-1,6x})^2}. \tag{5}$$

Функция распределения некоторой величины является ее исчерпывающей вероятностной характеристикой. Но она имеет недостаток, заключающийся в том, что по ней трудно судить о характере распределения этой величины в небольшой окрестности той или иной точки числовой оси. Более наглядное представление о характере распределения данной величины в окрестностях различных точек дается плотностью распределения.

Вспомним, что плотностью распределения f(x) некоторой величины X в точке x называется производная ее функции распределения в этой точке. График плотности распределения f(x) называют кривой распределения:

$$f(x) = F'(x) = \frac{d}{dx}F(x). \quad (6)$$

Пусть $f_1(x)$ – функция плотности распределения входных данных, а $f_2'(x)$ – первая производная от АФН. Предположив, что в процессе обучения в результате подстройки коэффициентов на первом слое ИНС происходит масштабирование и сдвиг АФН с целью наиболее точного восстановления данных, введем в рассмотрение:

$$f_3(x) = af_2'(x) + b, \quad (7)$$

где a и b – коэффициенты, минимизирующие отклонение:

$$K = \int_{-\infty}^{\infty} \left| f_1(x) - f_3(x) \right| dx. \quad (8)$$

Таким образом, введенная величина К в данном случае будет численно характеризовать несоответствие между АФН входного слоя ИНС и входными данными в отношении их статистических характеристик и может служить показателем чувствительности ИНС к входным данным (рис. 2).

Ниже приведено графическое пояснение изложенного выше: на рис. 2 изображена идеализированная ситуация, а на рис. 3 – реальная. Нетрудно заметить, что чем больше площадь несовпадения (заштрихована), определяемая показателем К, тем хуже чувствительность ИНС к данному виду распределения входных данных.

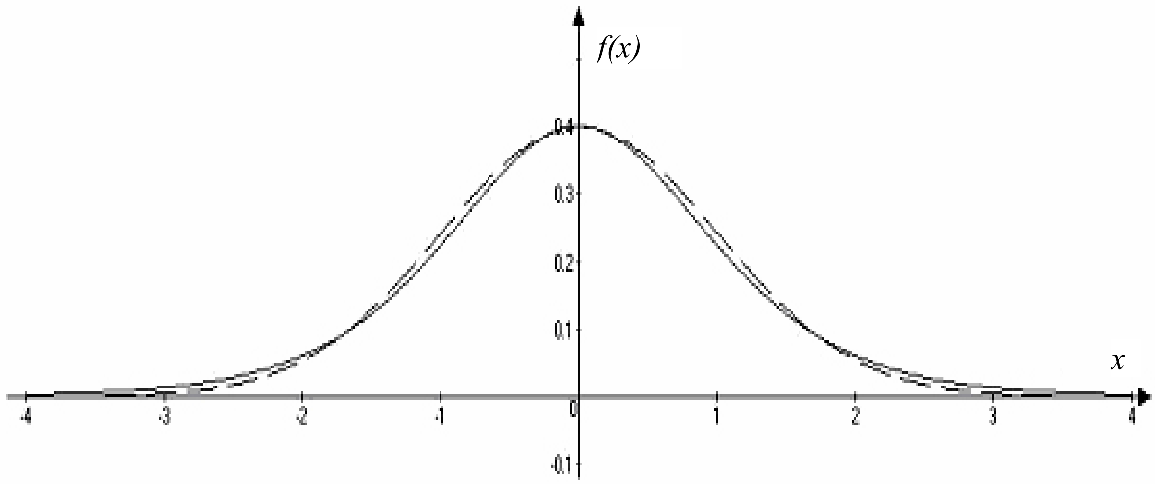


Рисунок 2 – График несовпадения нормированных функций:
$f_1(x)$ – кривая нормального распределения входных данных (пунктиром);
$f_3(x)$ – первая производная нормированной и отмасштабированной сигмоидальной АФН

Наименьшее значение показателя чувствительности К будет свидетельствовать о наибольшем совпадении (рис. 3 и рис. 4), а также о наибольшей чувствительности ИНС с данной АФН к значениям исследуемого временного ряда (ВР).

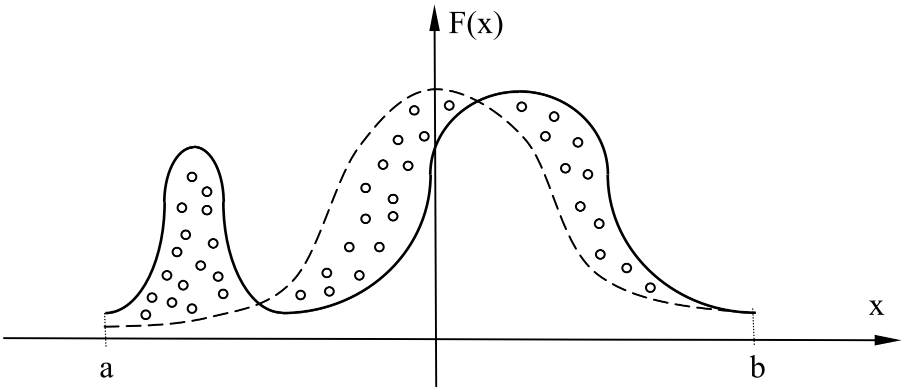


Рисунок 3 – График несовпадения функций: сложного вида распределения (полимодального) и сигмоидальной АФН (пунктиром)

Круг задач, решаемых ИНС

Активационная функция нейрона (АФН) во многом определяет круг задач, которые ИНС может решать. В настоящее время выбор типа АФН крайне не формализован и, как правило, подбор АФН и настройка ее параметров производится на основании определенных видений эксперта на базе эвристических и эмпирических предпосылок.

Используя предложенный в (8) показатель К, становится возможным довольно точно определить виды распределений входных данных, присущих тому или иному классу задач, наиболее успешно решаемых ИНС. Другими словами, теперь, зная величину показателя К, можно легко определить, насколько та или иная АФН подходит для обработки входных данных, имеющих известный закон распределения. Проведя данный анализ для каждой из основных АФН, можно для каждой из них определить класс задач, для которой они более всего подходят.

Из определения показателя К, очевидно, вытекает рекомендация о нежелательности использования унифицированной АФН одного вида (будь то сигмоидальная или иная функция) к различным входным данным, имеющим разные виды распределений, как это принято делать в настоящее время, поскольку у унифицированной АФН отсутствует возможность полноценно и гибко выполнять свои передаточные функции из-за игнорирования статистических характеристик входных данных. Наглядно данная рекомендация проиллюстрирована попыткой наложить сигмоидальную АФН на кривую распределения входных данных, имеющих сложный вид распределения (рис. 3).

Таким образом, в данной работе было показано, что, например, для входных данных, имеющих нормальный закон распределения, наиболее эффективной будет сигмоидальная АФН (рис. 2), в то время как для входных данных, имеющих равномерный закон распределения, наиболее эффективной будет линейная АФН.

На практике входные данные обычно представлены сложными видами распределений, в частности полимодальным, антимодальным и др. В любом случае необходимо подбирать такие АФН и так настраивать их параметры, чтобы показатель К стремился к нулю.

На основании идеи, изложенной выше, были составлены справочные материалы (табл. 1), включающие в себя данные о соответствии между различными видами распределений входных данных и основными типами АФН входного слоя ИНС, а также методика применения показателя К.

Таблица 1 – Типовые соответствия различных видов распределений входных данных и основных типов АФН входного слоя ИНС

№	Название распределения	Функция распределения	Название АФН	Функция АФН
1	Нормальное	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x\in R$	Сигмоидальная	$F(x)=\frac{1}{1+e^{-ax}}$
2	По закону Лапласа	$f(x)=be^{-\lvert x\rvert}$		
3	Равномерное	$f(x)=\begin{cases}\frac{1}{b-a}, x\in(a,b);\\ 0, x\notin(a,b)\end{cases}$	Линейная	1) $x\le a$ $F(x)=0$ 2) $x\in(a,b]$ $F(x)=\frac{x-a}{b-a}$ 3) $x>b$ $F(x)=1$
4	Показательное	$f(x)=\begin{cases}\lambda e^{-\lambda x}, x>0\\ 0, x\le 0\end{cases}$	Экспоненциальная	$(x>0)$ $F(x)=1-e^{-\lambda x}$

Методика применения показателя К

Методику применения разработанного показателя К в общем случае можно представить в виде последовательности этапов:

1. Исследование входных данных на предмет выявления их статистических характеристик.
2. Определение вида распределения этих данных.
3. Выбор одной из типовых ситуаций (табл. 1), т.е. определение наиболее подходящей АФН к данному виду входных данных.
4. Соответствующая настройка АФН входного слоя данной ИНС.

Приведенную методику следует применять в случаях, когда ИНС обрабатывает ВР, представленный значениями некоторой непрерывной величины.

К перспективам развития идеи, изложенной в данной работе, следует отнести доработку методики применения предложенного показателя для случая обработки дискретных величин, а также расширение базы данных справочных материалов в табл. 1.

Выводы

В данной работе рассмотрена проблема потери информативности входных данных на этапе синтеза ИНС. Для ее решения предложен подход, позволяющий подбором АФН и специальной настройкой её параметров учитывать статистические характеристики входных данных, что, в свою очередь, позволило ввести и обосновать показатель чувствительности ИНС к статистическим характеристикам входных данных.

Разработка такого показателя чувствительности дала возможность определять круг задач, которые ИНС может решать наиболее эффективно, используя ту или иную АФН, а в перспективе – разработку метода, позволяющего сохранять исходную информативность входных данных.

Литература

1. Саймон Хайкен. Нейронные сети: полный курс. – 2-е изд., исп.: Пер. с англ. – М.: ООО «Вильямс», 2006. – 1104 с.
2. Уоссерман Ф. Нейрокомпьютерная техника: теория и практика. – М.: Мир, 1992.
3. Сборник научных трудов V Всероссийской научно-технической конференции «Нейроинформатика – 2003» / Под ред. Мишулина О.А. – Ч. 1. – М.: МИФИ, 2003. – 244 с.
4. Востров Г.Н., Любченко В.В., Полякова М.В. Моделирование временных рядов с использованием вейвлет-сетей // Искусственный интеллект. – 2000. – № 3. – С. 207-214.
5. Крисилов В.А., Кондратюк А.В., Чумичкин К.В. Метод повышения чувствительности нейронной сети к исходным данным // Искусственный интеллект. – 2006. – № 4. – С. 290-296.
6. Миркес Е.М. Нейроинформатика: Учеб. пособие для студентов. – Красноярск: ИПЦ КГТУ, 2002.
7. Kondratiuk O., Krisilov V. Formalization of the structural synthesis of neural networks for the purposes of forecasting of time series // V[th] National Conference on Information Technologies. – Gdansk (Poland) ETI. – 2007. – Vol. 13.

А.В. Кондратюк, К.В. Чумичкін
Показник чутливості нейронної мережі до вхідних даних
У статті досліджена проблема втрати інформативності вхідних даних на етапі синтезу штучної нейронної мережі у вирішенні задач прогнозування часових рядів. Запропоновано підхід до вирішення даної проблеми, заснований на підвищенні чутливості нейронної мережі до вхідних даних шляхом аналізу їх статистичних характеристик. Запропоновано вирішення актуальної задачі розробки показника, що визначає ступінь чутливості нейронної мережі до значень вхідних даних. Завдяки розробці такого показника стало можливо визначити види вхідних даних, що найбільш ефективно обробляються нейронною мережею. Ефективність запропонованого показника обґрунтована і підтверджена теоретичними викладками та проілюстрована наглядно на рисунках. Визначені перспективи подальшого розвитку даного напрямку досліджень.

O. Kondratiuk, K. Chumychkin
Sensitivity Index of Neural Network to Input Data
The article is devoted to investigation of the problem of informativity dump input data at the stage of artificial neural network synthesis for solution of the task of time series prediction. The new approach to solve this task is offered which is based on neural network sensitivity to input data by analysis of their static characteristics. The new approach to solve actual task of development index which is defined the degree of sensitivity neural network to values of input data. It becomes possible to indicate the sort of input data owing to development such index which is processed the most effective by neural network. The efficiency of the offered index is justified and proved by theoretical calculations and illustrated obviously in the pictures. Perspective of further development of this direction research is indicated.

Статья поступила в редакцию 30.11.2007.